09/830497
Rec'd PCT/PTO 27 JUN 2001

Description

Method for identification of speakers on the basis of their voices

The invention relates to a method for identification of speakers on the basis of their voices.

The object on which the invention is based is to specify a method for identification of speakers on the basis of their voices, which method is robust, safe, secure and reliable.

According to the invention, this object is achieved by the features specified in patent claim 1.

The invention will be described in more detail in the following text using a flowchart.

1.
The invention allows the identification of the speaker on the basis of his voice. The problem of speaker identification is to distinguish between different speakers or to check the predetermined speaker identity, with the only input information being the recording of the voice of the speaker.

Furthermore, a method is proposed which prevents the access system from being outwitted when the voice and the keyword are recorded by third parties.

When complex probability distributions for the speech parameters of a speaker are stored, a compromise must be made between accuracy and memory requirement. For this reason, methods for storage of the probability distributions have been proposed which can be used as a function of the number of speakers.

2.
Until now, the speaker has been identified, for example, with the aid of hidden Markov models or by vector quantization, see reference [1].

3.
The invention solves the problem of speaker identification based on the parameters of an analysis by means of synthesis coders using linear prediction (LPAS) [1] (for example a harmonic vector excited codec [5] or waveform interpolation codec [4]. The speech signal parameters used in the past, such as Cepstral AR parameters, do not provide a satisfactory solution to the problem. For this reason, other parameters need to be used, such as parameters relating to the excitation of the vocal tract, which include information that is dependent on the speaker and is at the same time largely phoneme-independent.

Furthermore, the method for estimation of the probability distribution of the coder parameters for the respective speaker is given, as well as a method which prevents the access system from being outwitted.

Speaker identification

In systems for *speaker identification*, statistical principles [2] are used to check whether the spoken sentence has been spoken by one of the speakers covered by the speaker identification system. In the process, there are in principle two types of speaker identification systems, text-dependent systems and text-independent systems. For the procedure described in the invention, text independence of the system is achieved by means of an expanded training phase, in which the speaker has to record a wide range of material, and the probability distributions of said speech signal parameters are established from all the

spoken material. A text-dependent system can be trained more easily since the spoken material which is spoken by the speaker during the usage phase is limited to a number of keywords or specific

sentences. The preparation phase is continued until the system reliably identifies the voice of the speaker. The object of *speaker identification* is illustrated in Figure 2.

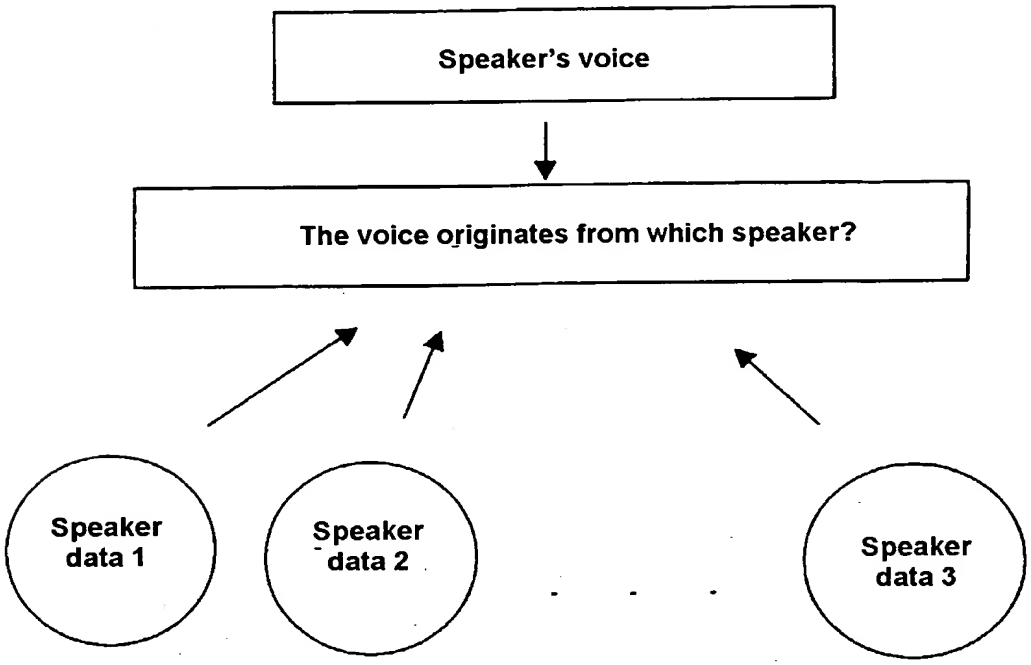

Figure 2. The problem of speaker identification

Speaker identification is dealt with as a problem relating to the detection of multiples [2]. The classes to be distinguished between, one for each speaker who is intended to be identified by the system, are referred to as sp_i = 1..M, where M is the number of speakers covered by the speaker identification system. Speaker identification is based on recorded signals spoken by the respective speaker. The speech signal is segmented into the signal frames x = [$x(1)..x(K)$] (K = 160 for a signal frame with a length of 20 ms and a sampling frequency of 8 kHz, for example). The segmentation process produces the speech signal frames x(1)..x(N), where N depends on the total length of the sentence or keyword spoken by the speaker. The decision on the speaker is made from the probabilities or probability densities (referred to jointly as probability scores) that the vectors of the samples x(l) l = 1..N belong to the class sp_i. The statistically optimum decision scheme selects that class sp_i having the highest probability value for given x(l), l = 1..N. This means that the vector x(l) is assigned to the class sp_j for which:

$$p(\mathbf{x}(1)...\mathbf{x}(N) \mid sp_j) > p(\mathbf{x}(1)...\mathbf{x}(N) \mid sp_i) \quad \text{for all} \quad j \neq i$$

Speaker verification

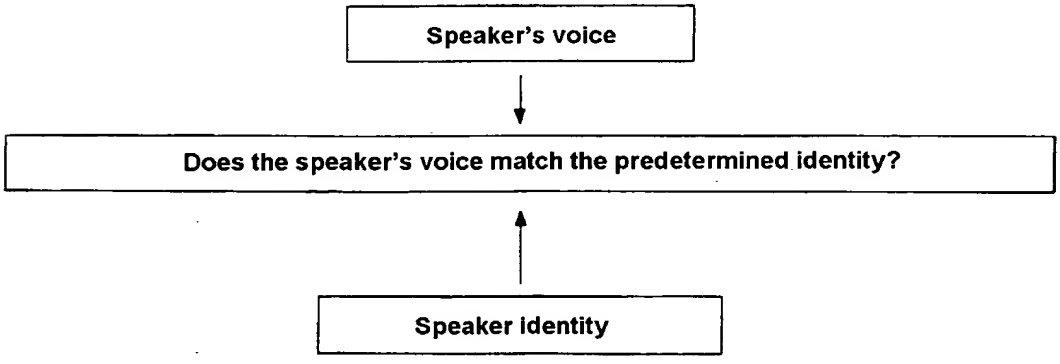

Figure 3. The problem of speaker verification

The problem of *speaker verification* is to check the predetermined identity of the speaker on the basis of his voice. This corresponds to the situation illustrated in Figure 3.

The process of speaker verification is carried out in a similar manner to that of speaker identification, that is to say the spoken sentence is likewise segmented. However, after this, the voice is not classified, but a probability score is calculated for the predetermined speaker identity, and is compared with a threshold. The identity of the speaker is thus confirmed on the basis of his voice when:

$$p(x(1)..x(N) \mid sp_j) > \text{threshold}$$

where sp_j corresponds to the predetermined speaker identity. The threshold must be set sufficiently high to avoid the situation in which a speaker with a different identity to that predetermined is accepted/authorized.

LPAS coder

The speech coding methods used nowadays are predominantly based on the analysis by synthesis method using an LPC synthesis filter [2]. In these methods, speech coding is optimized by repetition of the coding and decoding operations until the optimum parameter set is found for the given speech section.

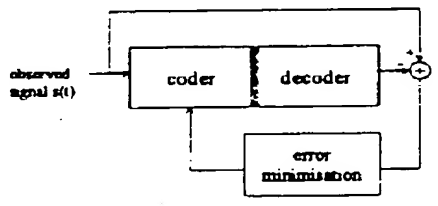

Figure 4: Design of an LPAS coder

One of the most widely used types of LPAS coder is the CELP coder. One relatively new development is the harmonic vector excited codec, where the form of the excitation signals is particularly suitable for the described task. Figure 4 shows the synthesis model of a CELP coder. The synthesis model defines the method for calculating the synthesized speech signal from the quantized parameters of the speech signal. In general, each LPAS coder has the following parameter groups:

- Short-term predictor parameters. The short-term predictor parameters are generally calculated by means of classical LPC analysis, using the correlation method or the covariance method for linear prediction [3]. 8-10 LPC coefficients are used for signal frames with a length of 20 to 30 ms and a sampling rate of 8 kHz. The short-term predictor parameters may occur in various forms (for example the reflection coefficients or in the form of line spectrum frequencies LSF), depending on the representation which can be quantized better. It has been found that the LSF coefficients are most suitable for quantization, and this form of prediction coefficients is generally used. The short-term predictor parameters are calculated using an open-loop procedure, that is to say without the overall optimization, illustrated in Figure 1, with the other parameters relating to the synthesis error.

- Long-term predictor parameters. Long-term predictor parameters are used in a filter which synthesizes the fundamental frequency of the speech signal. This is generally a long-term predictor with a filter coefficient and a parameter for the fundamental period of the voice signal. A long-term predictor with the parameters b = [b,N] is a part of Figure 2. The long-term predictor parameters are likewise calculated using an open-loop procedure, without

overall optimization with the other parameters. In some

coders, a refined search is sometimes carried out based on the long-term predictor parameters using a closed-loop procedure.

- The excitation parameters. The 5-10 ms subframes of the remaining signal are vector-quantized using a closed-loop procedure in a CELP coder. The transmitted parameters allow the signal forms to be reproduced at the decoder end from the stored codebook.

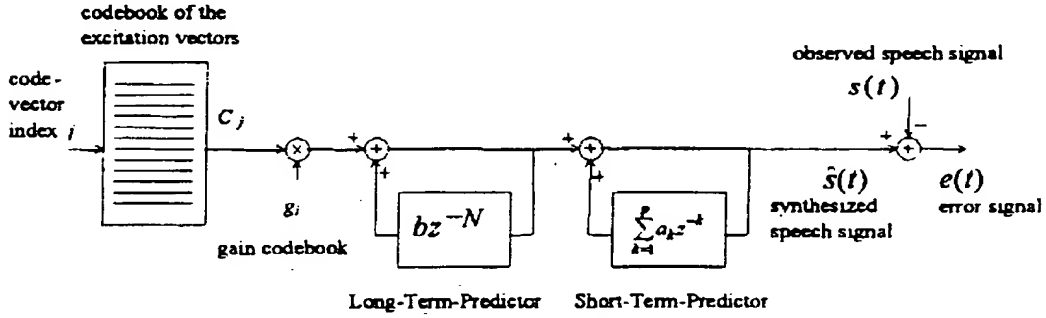

Figure 5: The synthesis model for a CELP coder

In an HVXC codec, the output from the LPC analysis filter is transformed to the frequency domain and the fundamental-period-normalized spectral envelope is vector-quantized.

Speaker identification using the parameters of an LPAS coder

The speech coder parameters provide a comprehensive description of the possible speech signals using considerably fewer parameters than when the speech signal is represented as a sequence of samples.

The decomposition of the speech signal into the said parameter groups can be used in various ways for speaker identification. The methods for calculation of the parameters and synthesis of the speech signal imply probability density estimation methods (for example the

probabilities of the parameters, which are regarded as discrete probability variables). Those defined using a closed-loop procedure should actually be

regarded as discrete probability variables, since it is impossible to link the volumes of the parameter space regions of the vector quantizer for parameters such as these. This relates in particular to the excitation parameters. The probability distributions for such parameters are estimated by calculating relative frequencies of the parameters/code vectors in the training statement.

Those which are calculated using an open-loop procedure in the coder are initially available in a non-quantized form and are quantized only after this, with vector quantization generally being used. For parameters such as these, the probability densities can be estimated from the training statement. This approach is used primarily for the short-term predictor parameters.

The probability density estimation is based on the histogram method [6]. This method requires knowledge of the volumes of those regions of the parameter space which are linked to the quantized points.

A method for storage of probability distributions is obtained if the possible code vectors for the speech signal parameters are stored once for the entire population, which corresponds to the situation where the quantization steps/code vectors are determined once, from the database which contains the recordings by a large number of speakers. The probability distributions of the parameters for the speakers are then stored together with the indices of the code vectors for the parameters in the system. This is suitable for large systems with a very large number of users (ATM, access systems in companies).

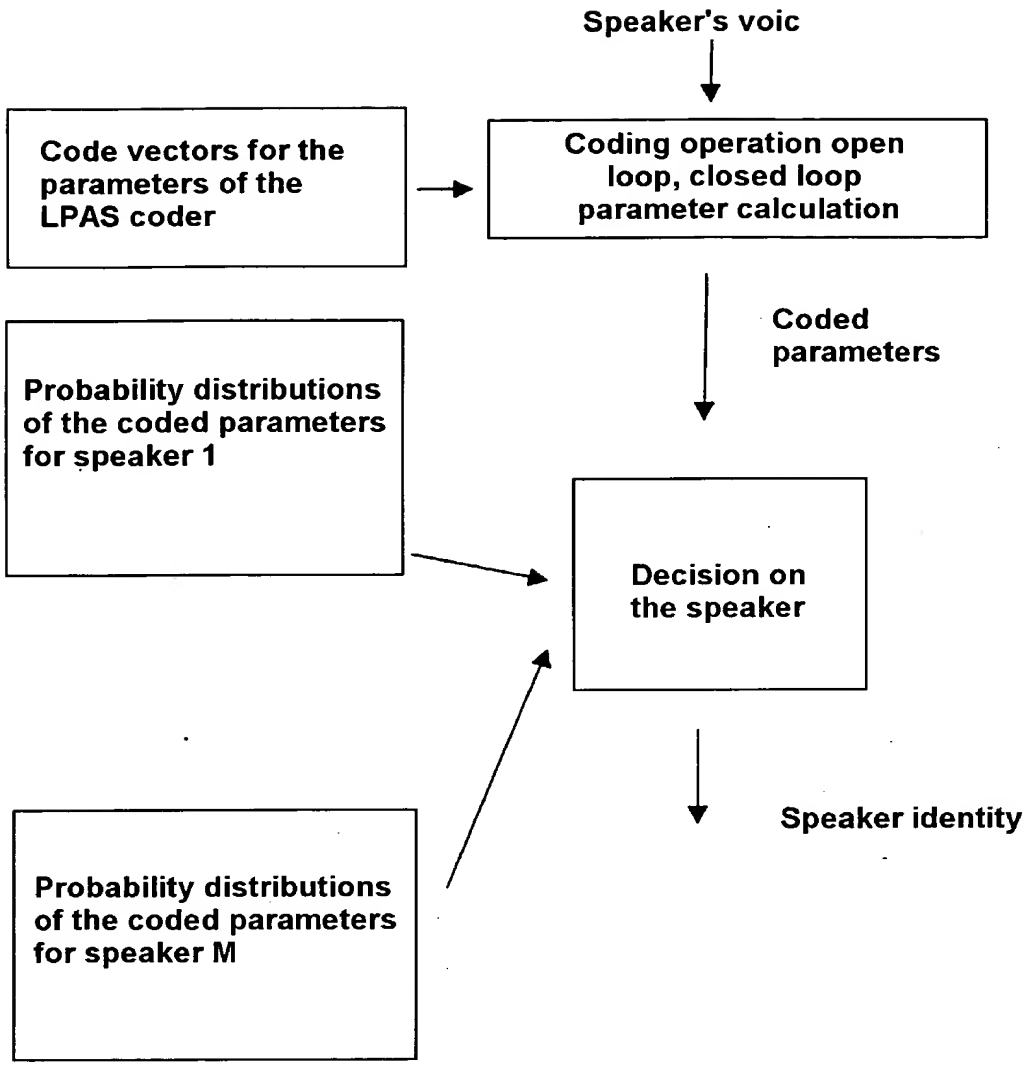

Figure 6. Speaker identification using the parameters of an LPAS coder

Another method is for the code vectors for the parameters for each speaker to be trained individually. The code vectors are then stored together with the values of the probability densities at those parameter space points defined by the code vectors. One possible way of carrying out this method is shown in Figure 7. This method is intended for a small number of speakers (for example for a voice-controlled door in a dwelling).

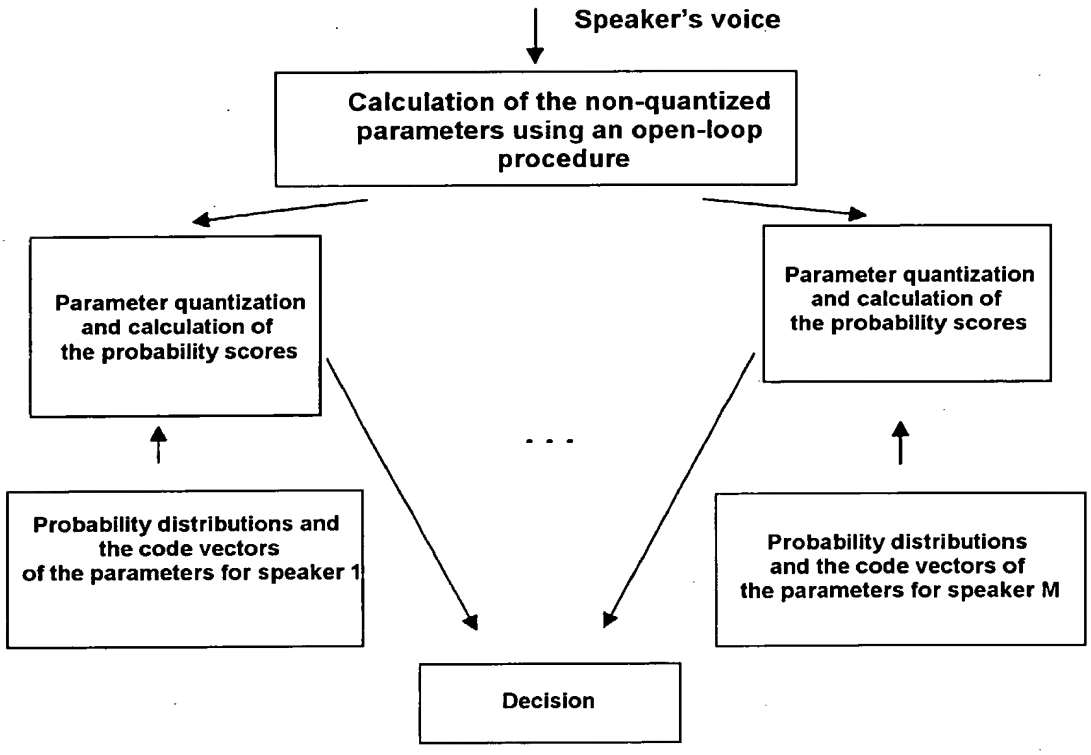

Figure 7. Speaker identification using the parameters of an LPAS coder

Probability densities are stored together with the code vectors for the parameters

↓ **Speaker identity**

Training phase of a speaker identification system

The probability density distributions for the speaker classes are estimated from the training material. For text-dependent speaker identification (speaker identification/speaker verification), a specific sentence or keyword is repeated during the training phase until the speaker identification operates reliably.

Phonetically balanced spoken material must be recorded for text-independent speaker verification. In this case as well, the training phase must be repeated until the speaker identification/verification operates reliably.

The material recorded during the training phase is in each case used with a phase shift a number of times for training, in order to make the speaker identification system independent of the initial phase of the recorded voices. The data used for training are referred to as the training statement TS_{sp_i}, with sp_i symbolizing the speaker.

Estimation of the probability densities

In order to describe the method according to the invention for estimation of the probability densities of the parameters for the speaker classes, a number of necessary definitions will be introduced first of all.

The introduced abstraction of the coding process has the advantage that the estimation of the probability densities can be described in a simple manner without needing to go into details of the highly complicated operations in the speech coder. A detailed description of the parameter calculation process can be found in [4] and [5].

A speech coder operates in evaluation intervals. The operations described in that section via the LPAS coder are carried out in the speech coder for each signal frame, and

supply the parameters of the speech signal for the respective frame.

Calculation of a non-quantized parameter vector p from the signal frame x is written as $p = K_p(x)$ in an open-loop optimization procedure. The quantization of the parameter is referred to as $\hat{p} = Q_p(p$. That region in the parameter space of the parameter p which is mapped onto the code vector $\hat{p}$ in the coding process is referred to as $S_{\hat{p}} = \{p : Q_p(p) = \hat{p}$. The volume of this region is referred to as $V(S_{\hat{p}}$.

The set of possible code vectors for the parameter p is written as $C_p = \{\hat{p}; i = 1..N_p\}$, where N_p is the number of code vectors. The set or regions which are linked to the code vectors is referred to as $R_p = \{S_i; i = 1...N_p$.

The association function for a region S_i is referred to as:

$$1_{S_i}(\mathbf{p}) = \begin{cases} 1 \text{ for } \mathbf{p} \in S_i \\ 0 \text{ for } \mathbf{p} \notin S_i \end{cases}$$

The frequency of occurrence of a parameter in the training statement is calculated using

$$f_{S_i} = \frac{\text{Number of parameter values from the training statement } TS_{sp_i} \text{ which occur in the region } S_i}{\text{Number of parameter values from the training statement } TS_{sp_i}}$$

The estimated probability density distribution then becomes:

$$p(\mathbf{p} \mid sp_i) = \sum_{k=1}^{N_p} 1_{S_k}(\mathbf{p}) \frac{f_{S_i}}{V(S_i)}$$

Estimation of the probabilities

The probability functions (probability mass functions) are estimated for those parameters which are regarded as discrete probability variables, that is to say in particular the excitation from the codebook, which is

optimized using a closed-loop procedure, and the fundamental period of the speech signal. These probability functions are defined as the frequencies of the given

parameter codes in the training statement for the respective speaker.

Storage of the probability distributions

The speech parameters are not all calculated at the same time, but successively, in a speech coder. For example, the short-term predictor parameters are calculated first, and the remaining parameters for already known short-term predictor parameters are then optimized with regard to the synthesis or the prediction error. This allows effective storage of the probability distributions as conditional probabilities of the code vectors in a tree structure. This is possible thanks to the following relationship:

$$p(\mathbf{p}_K, \mathbf{p}_L, \mathbf{p}_A \mid sp_i) = p(\mathbf{p}_K \mid sp_i)\, p(\mathbf{p}_L \mid sp_i, \mathbf{p}_K)\, p(\mathbf{p}_A \mid sp_i, \mathbf{p}_K, \mathbf{p}_L)$$

P_K - Vector for a short-term parameter
P_L - Vector for a long-term parameter
P_A - Vector for an excitation parameter

A major simplification can be achieved if the speech parameters within a signal frame can be assumed to be statistically independent. The above formula then becomes:

$$p(\mathbf{p}_K, \mathbf{p}_L, \mathbf{p}_A \mid sp_i) = p(\mathbf{p}_K \mid sp_i)\, p(\mathbf{p}_L \mid sp_i)\, p(\mathbf{p}_A \mid sp)$$

The probability densities need to be stored at a very large number of points in parameter space in the system. The number of bits used for storing probability densities is critical to the complexity of the overall system. A vector quantizer is therefore used for the probability values. This makes it possible to reduce the number of bits used for storing the probability distributions.

System safety and security

In order to prevent the system from being outwitted, noise is transmitted at the same time that the voice of the speaker is being recorded, which noise is known to the system, and from which the digitized speech signal is subtracted.

5.

The invention can be used for access control applications, such as voice-controlled doors, or for verification, for example for bank access systems. The procedure can be implemented as a program module on a processor which carries out the task of speaker identification in the system.

[1] S. Furui, "Recent advances in speaker recognition", Pattern Recognition Letters, Tokyo Inst. of Technol., 1997

[2] P. Vary, U. Heute, W. Hess, *Digitale Sprachsignalverarbeitung* [*Digital speech signal processing*], B.G. Teubner, Stuttgart, 1998

[3] K. Kroschel, *Statistische Nachrichtentheorie* [*Statistical information theory*], 3rd ed., Springer-Verlag, 1997

[4] W.B. Kleijn, K.K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995

[5] ISO/IEC 14496-3, MPGA-3 HVXC Speech Coder description

[6] Prakasa Rao, Functional Estimation, Academic Press, 1982

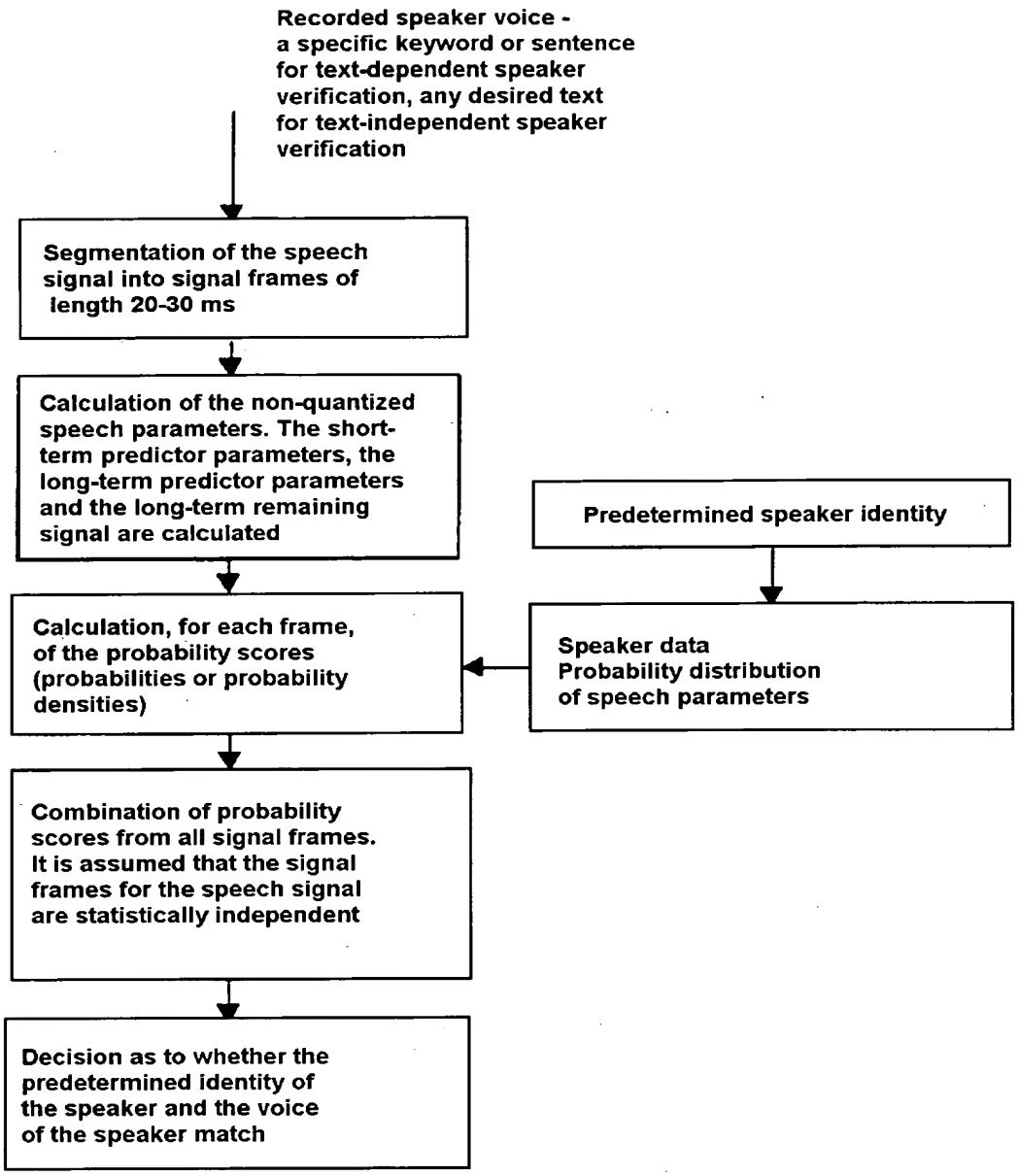

Figure 1. Speaker verfication using the parameters of an LPAS coder

Patent Claims

1. A method for identification of speakers on the basis of their voices, having the following features:

(a) in a preparation phase,

(a1) k text-dependent or text-independent reference spoken expressions, which form a speaker-related training statement, from M speakers are segmented into first speech signal frames of length L,

(a2) the first speech signal frames are supplied to an analysis-by-synthesis coder based on linear prediction,

(a3) a first short-term predictor parameter, long-term predictor parameter and/or excitation parameter for the coder are/is calculated in the analysis-by-synthesis coder for each of the M speakers and for each first speech signal frame in each case, with the parameters then forming speaker-related training material,

(a4) the frequency of the respective occurrence of the first short-term predictor parameter, of the long-term predictor parameter and/or of the excitation parameter for the coder in the speaker-related training statement and/or the probability densities with which the first short-term predictor parameter, the long-term predictor parameter and/or the excitation parameter are/is contained in the speaker-related training statement are/is calculated in the analysis-by-synthesis coder for each of the M speakers and for each first speech signal frame in each case,

(a5) the calculated frequencies and/or probability densities are stored on a speaker-related basis as speaker data,

(b) in a simulated usage phase of the training phase,

(b1) a text-dependent or text-independent simulation spoken expression of an m-th speaker where m=1..M is segmented into second speech signal frames of length L,

(b2) the second speech signal frames are supplied to the analysis-by-synthesis coder,

(b3) a second short-term predictor parameter, long-term predictor parameter and/or excitation parameter for the coder are/is calculated in the analysis-by-synthesis coder for the m-th speaker and for every other speech signal frame in each case,

(b4) first probability hits are calculated for every other speech signal frame from the calculated second short-term predictor parameter, long-term predictor parameter and/or excitation parameter and the speaker data stored for the m-th speaker in the preparation phase, which probability hits indicate the probability with which the second short-term predictor parameter, long-term predictor parameter and/or excitation parameter match(es) the first short-term predictor parameter, long-term predictor parameter and/or excitation parameter,

(b5) the first probability scores from all the second speech signal frames are combined,

(b6) a check is carried out to determine whether the combined first probability scores are greater than a predetermined first threshold which confirms the voice of the m-th speaker, when the combined first probability scores are greater than the predetermined first threshold or the preparation phase continues for a further i reference spoken expressions by the m-th speaker until the voice of the m-th speaker is confirmed, when the combined first probability scores are less than or equal to, or are less than, the predetermined first threshold,

(c) in a usage phase

(c1) a text-dependent or text-independent used spoken expression of the m-th speaker where m=1..M is segmented into third speech signal frames of length L,

(c2) the third speech signal frames are supplied to the analysis-by-synthesis coder,

(c3) a third short-term predictor parameter, long-term predictor parameter and/or excitation parameter for the coder are/is calculated in the analysis-by-synthesis coder for the m-th speaker

and for every third speech signal frame in each case,

(c4) second probability hits are calculated for every third speech signal frame from the calculated third short-term predictor parameter, long-term predictor parameter and/or excitation parameter and the speaker data stored for the

m-th speaker in the preparation phase, which second probability hits indicate the probability with which the third short-term predictor parameter, long-term predictor parameter and/or excitation parameter have been spoken by the m-th speaker,

(c5) the second probability hits from all the third speech signal frames are combined,

(c6) a check is carried out to determine whether the combined second probability scores are greater than a predetermined second threshold and the voice of the m-th speaker is identified when the combined second probability hits are greater than the predetermined second threshold, or the voice of the m-th speaker is not identified when the combined second probability scores are less than or equal to, or are less than, the predetermined second threshold.

2. The method as claimed in claim 1, characterized in that
a harmonic vector excited predictive coder or a waveform interpolating coder is used, in particular, as a parametric coder.

3. The method as claimed in claim 1, characterized in that
a coder based on linear prediction, in particular an LPAS coder, is used as the analysis-by-synthesis coder.

4. The method as claimed in one of claims 1 to 3, characterized in that
the frequencies and/or probability densities are quantized using a vector quantizer having a specific, considerably reduced, number of bits.

5. The method as claimed in one of claims 1 to 4, characterized in that

noise which is known to the speaker identification system is also entered when the spoken expression of the speaker is entered into the speaker identification system.

6. The method as claimed in one of claims 1 to 5, characterized in that the noise which is also entered is subtracted internally, before the segmentation, from the recording of the speaker voice.

1999P02665

Abstract

Method for identification of speakers on the basis of their voices

The invention relates to a method for speaker identification using parameters of an LPAS coder or of a parametric coder for modeling the probability distribution for the speaker classes.

FIGURE CAPTIONS

1 **Speaker identification system preparation phase*** (Profile for speaker j)

2 Training of a text-independent system

3 Training of a text-dependent system

4 Recording of widely phonetically balanced material from the j-th, j=1..M system user. A relative large number 1..K of reference spoken expressions.

5 Specific word sequence, a sentence or a key word. Corresponding number 1..K of reference spoken expressions for the j-th, j=1..M system user.

6 Segmentation of the training material into signal frames x(1)...x(N) where N is dependent on the total length of the spoken expressions. x(i)=[x(1)...x(L) where L is the length of the signal frame.

7 Large number of speakers >10

8 Speech database several hours of recordings by different speakers

9 Training of the speaker-independent codebooks for the short-term parameters with the aid of the K-mean algorithm $Cb_K = [C_{Ki} \in R^p, i=1..L_K]$, L_K is the number of codebook entries and p = 8..10 length of the LSF code vector.

10 Training of the speaker-independent codebooks for the excitation parameters. Codebooks for the fundamental-period-normalized spectral forms of the LPC remaining signal $Cb_K = [C_{Ai} \in R^p, i=1..L_A]$,

(L_A is the number of codebook vectors and p = 44 length of the code vector). Parameters in the same form as in the HVXC codec.

11 * The process defined from hereon is carried out for each new user of the speaker identification system. The aim of the preparation phase is to produce the speaker data for each of the M speakers.

12 Small number of speakers <10

13 The codebooks are trained for each of the M speakers using the material recorded by the respective speaker

14 Training of the speaker-independent codebooks for the short-term parameters with the aid of the K-mean algorithm $Cb_K = [C_{Ki} \in R^P, i=1..N_K]$, N_K is the number of codebook entries and p = 8..10 length of the LSF code vector.

15 Training of the speaker-independent codebooks for the excitation parameters. Codebooks for the fundamental-period-normalized spectral forms of the LPC remaining signal $Cb_K = [C_{Ai} \in R^P, i=1..N_A]$, ($N_A$ is the number of code vectors and p = 44 length of the harmonic code vector). Parameters in the same form as in the HVXC codec.

16 Calculation of the speech parameters for the training sets for each speaker based on the layout of an HVXC codec*

17 Calculation of the short-term parameters for each signal frame $p_K(i)$, i = 1..N

Training set for the short-term parameters is formed for the respective speaker: $TSK_j = \{p_K(i), i = 1..N\}$ $j=1..M$

18 Calculation of the long-term parameters for each signal frame $p_L(i)$, $i = 1..N$
Training set for the long-term parameters is formed for the respective speaker: $TSL_j = \{p_L(i), i = 1..N\}$

19 Calculation of the excitation parameters for each signal frame $p_A(i)$, $i = 1..N$
Speech fundamental-period-normalized spectral forms of the LPC remaining signal
Training set for the short-term parameters is formed for the respective speaker: $TSA_j = \{p_A(i), i = 1..N\}$

20 * ISO/IEC 14496-3 Information Technology - Very low bit rate audio-visual coding

21 Calculation of the volumes of the Voronoi cell regions for the probability density estimate of the short-term predictor parameters

22 Calculation of the volumes of the vector quantizer cells $S_{Ki} = \{X \in R^P: |C_{Ki} - x| < |C_{Kj} - x|, i \# j\}$

23 Calculation of the frequencies of the short-term parameters

$$f_{C_{Ki}} = \frac{\text{Number of code vectors } C_{Ki} \text{ contained in the } TSK_j}{\text{Number of all code vectors in the } TSK_j}$$

24 Calculation of the probability densities of the short-term predictor parameters:

25 ***Speaker$_j$***

26 Number of code vectors in the codebook Cb_K

27 $$1_{S_{Ki}}(p_K)=\begin{cases}1\ for\ p\in S_{Ki}\\0\ for\ p\notin S_{Ki}\end{cases}$$

28 Association function for the region S_{Ki}

29 Store the probabilities of the short-term predictor parameters for a large number of speakers

Code vector index 1 Probability density value 1
Code vector index j Probability density value 1

J - Number of Voronoi cells whose probability is not equal to zero.

30 Store the probabilities of the short-term predictor parameters for a small number of speakers

Code vector 1 Probability density value 1
Code vector I Probability density value I

I - Number of code vectors in the codebook of the short-term predictor parameters for speaker j.

31 Calculation of the frequencies of the long-term predictor parameters for the speaker j in the training set TSL_j

32 Speakers with the probability distributions of the long-term predictor parameters. These probability distributions are stored in the same way, irrespective of the number of speakers

Speech fundamental period value 1 Frequency 1
Speech fundamental period value D Frequency D

33 Calculation of the frequencies of the excitation parameters for speaker j in the training set TSA_j

34 Speakers with the probabilities of the excitation parameters for a large number of speakers

Code vector index 1 Probability value 1
Code vector index D Probability value D

D - Number of excitation code vectors whose probability is not equal to zero.

35 Speakers with the probabilities of the excitation parameters for a small number of speakers

Code vector index 1 Probability value 1
Code vector index L_A Probability value L_A

L_A - Number of code vectors in the codebook for the short-term predictor parameters for speaker j

36 Simulated usage phase
Training of the system for speaker j

37 Request to record the K+1 test spoken expression

38 Simulated usage phase for a text-independent system

39 Simulated usage phase for a text-independent system

40 Recording of any desired K+1 spoken expression by the j-th system user

41 Specific word sequence, a sentence or key word.
K+1-th spoken expression by the j-th system user

42 Segmentation of the test spoken expression into the signal frames x(1)...x(N) where N is dependent on the total length of the test expression. x(i)=[x(1)...x(L)] where L is the length of the signal frame.

43 Calculation of the speech parameters for the test spoken expression

44 Calculation for the short-term parameters $p_k(i)$, i = 1..N in each frame of the probability $p(p_k(i)|\text{speaker j})$

45 Calculation for the short-term parameters $p_k(i)$, i = 1..N in each frame of the probability $p(p_L(i)|\text{speaker j})$

46 Calculation for the short-term parameters $p_A(i)$, i = 1..N in each frame of the probability $p(p_A(i)|\text{speaker j})$

47 Calculation of the probability scores for each signal frame:
$p(p_K(i), p_L(i), p_A(i)|\text{speaker j}) = p_K(i)p_L(i)p_A(i)$

48 Combination of the results from all signal frames.

Calculation of the test statistics

$$WS=\prod_{i=}^{N} p(p_K(i), p_A(i), p_L(i)|speaker\ j)$$

49 WS > Threshold

50 Repetition of the preparation phase of the speaker identification system

51 NO

52 YES

53 No additional training of the probability distributions required. The probability distributions are stored in the system and are ready for the usage phase.

54 **Speaker identification system usage phase**
(Profile for the speaker j)

55 Text-independent system

56 Text-dependent system

57 Recording of any desired spoken expression

58 Specific word sequence, a sentence or a key word (for example the name of the user).

59 Segmentation of the spoken expression into the signal frames x(1)...x(N) where N is dependent on the total length of the spoken expression. x(i)=[x(1)...x(L)] where L is the length of the signal frame.

60 Calculation of the speech parameters for the spoken expression

61 Probability distributions of the speech parameters or speaker l (in the form dependent on the number of system users)

Probability distributions for the short-term parameters

Probability distributions for the long-term parameters

Probability distributions for the excitation parameters

62 Probability distributions of the speech parameters or speaker j (in the form dependent on the number of system users)

Probability distributions for the short-term parameters

Probability distributions for the long-term parameters

Probability distributions for the excitation parameters

63 up to speaker M

64 Predetermined identity of the speaker

65 Calculation for the short-term parameters $p_K(i)$, i = 1..N in each frame of the probability $p(p_K(i) | \text{speaker j})$

66 Calculation for the long-term parameters $p_L(i)$, i = 1..N in each frame of the probability $p(p_L(i) | \text{speaker j})$

67 Calculation for the excitation parameters $p_A(i)$, i = 1..N in each frame of the probability $p(p_A(i) | \text{speaker j})$

68 Calculation of the probability scores for each signal frame: $p(p_K(i), p_L(i), p_A(i) | \text{speaker j}) = p_K(i)p_L(i)p_A(i)$

69 Combination of the results from all signal frames. Calculation of the test statistics $WS = \prod_{i=}^{N} p(p_K(i), p_A(i), p_L(i) | speaker\ j)$

70 Rejection

71 Confirmation of the speaker identity

72 Speaker identification

73 Calculation for the short-term parameters $p_K(i)$, i = l..N in each frame of the probability $p(p_K(i) | \text{speaker m})$, m=1..M

74 Results for each of the M speakers

75 Calculation for the short-term parameters $p_K(i)$, i = l..N in each frame of the probability $p(p_L(i) | \text{speaker m})$, m=1..M

76 Calculation for the short-term parameters $p_A(i)$, i = l..N in each frame of the probability $p(p_A(i) | \text{speaker m})$, m=1..M

77 Calculation of the probability scores for each signal frame: $p(p_K(i), p_L(i), p_A(i) | \text{speaker m}) = p_K(i) p_L(i) p_A(i)$ for each of the M speaker m=1..M

78 Combination of the results from all signal frames. Calculation of the test statistics $WS = \prod_{i=}^{N} p(p_K(i), p_A(i), p_L(i) | speaker\ m)$ for each of the M speaker m=1..M

79 Definition of the speaker identity.
Speaker j is chosen for home WS(j)>WS(i), j≠i

1999P02665

80 Speaker identity